



The University of Texas at Austin
Hildebrand Department of Petroleum
and Geosystems Engineering
Cockrell School of Engineering

2023

ENERGY AI HACKATHON

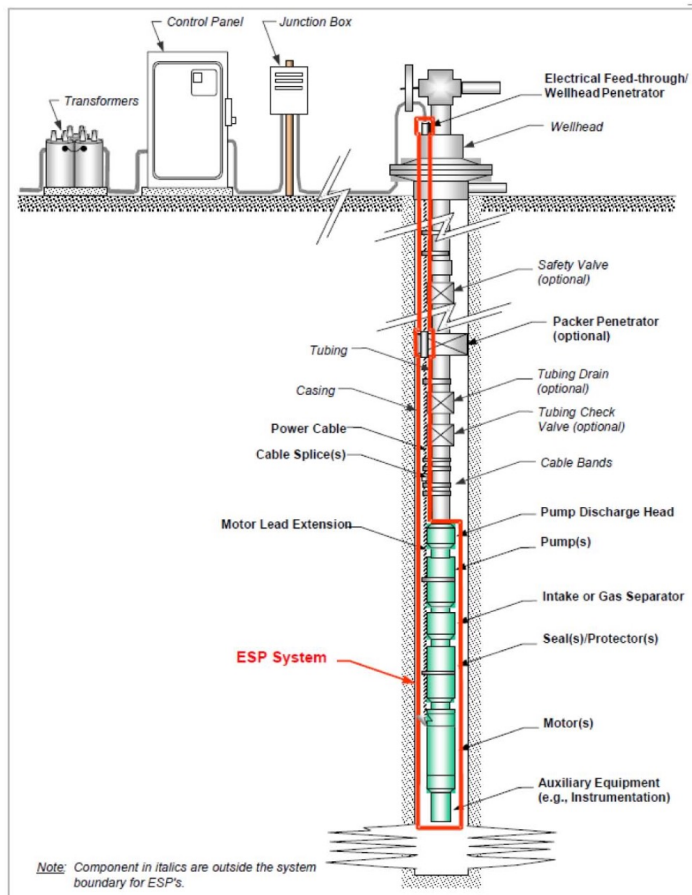
NFL - NO FREE LUNCH

Bernard Chang
Javier Guerrero
Martin Krylov
Olin Carty



PROBLEM DESCRIPTION

Prediction / classification "fail" or "not fail" within 30 days for 40 artificial lift
Electronic Submersible Pumps



Source: Abdelaziz et al. 2017



SCADA



Daily

Dynamic Data:

• *ESP Data*

- Discharge Pressure:
- Suction Pressure:
- Discharge Temperature:
- Motor Temperature:
- Vibration
- Motor Current:
- Leakage Current:
- Speed (Frequency)

• *Wellhead Data*

- Wellhead Temperature:
- Wellhead Pressure

• *Production Data*

- Well flow rate
- Water cut and GOR
- BS&W

Static Data:

• *Well Completion Data*

• *ESP Pump Design Information*

• *Reservoir Fluid Properties*

Historical Data:

• *Historical Well Events*

• *Investigation Reports*

• *Failure and Pull Reports*

• *DIFA Reports*

Source: Abdelaziz et al. 2017



DATA PREPARATION

Exploratory
Data Analysis



Feature
Imputation



Feature
Ranking

```
: # Drop columns with STR data or useless information
well_df = well_df.drop(['Unnamed: 0', 'Artificial_Lift_Type'], axis = 1)
well_solution_df = well_solution_df.drop(['Unnamed: 0', 'Artificial_Lift_Type'], axis = 1)

daily_df = daily_df.drop(['Lower_Limit', 'Power_Difference', 'Oil_Intake', 'Water_Intake', 'Gas_Saturation_at_Intake',
                          'Gas_through_Annulus_Intake', 'Gas_through_ESP_Intake', 'Gas_through_Annulus',
                          'Gas_Saturation_at_Discharge'], axis = 1)

daily_solution_df = daily_solution_df.drop(['Lower_Limit', 'Power_Difference', 'Oil_Intake', 'Water_Intake',
                                             'Gas_Saturation_at_Intake', 'Gas_through_Annulus_Intake',
                                             'Gas_through_ESP_Intake', 'Gas_through_Annulus',
                                             'Gas_Saturation_at_Discharge'], axis = 1)
```

Drop Features that can be reproduced by a combination of others features → **Avoid Collinearity**



Exploratory
Data Analysis

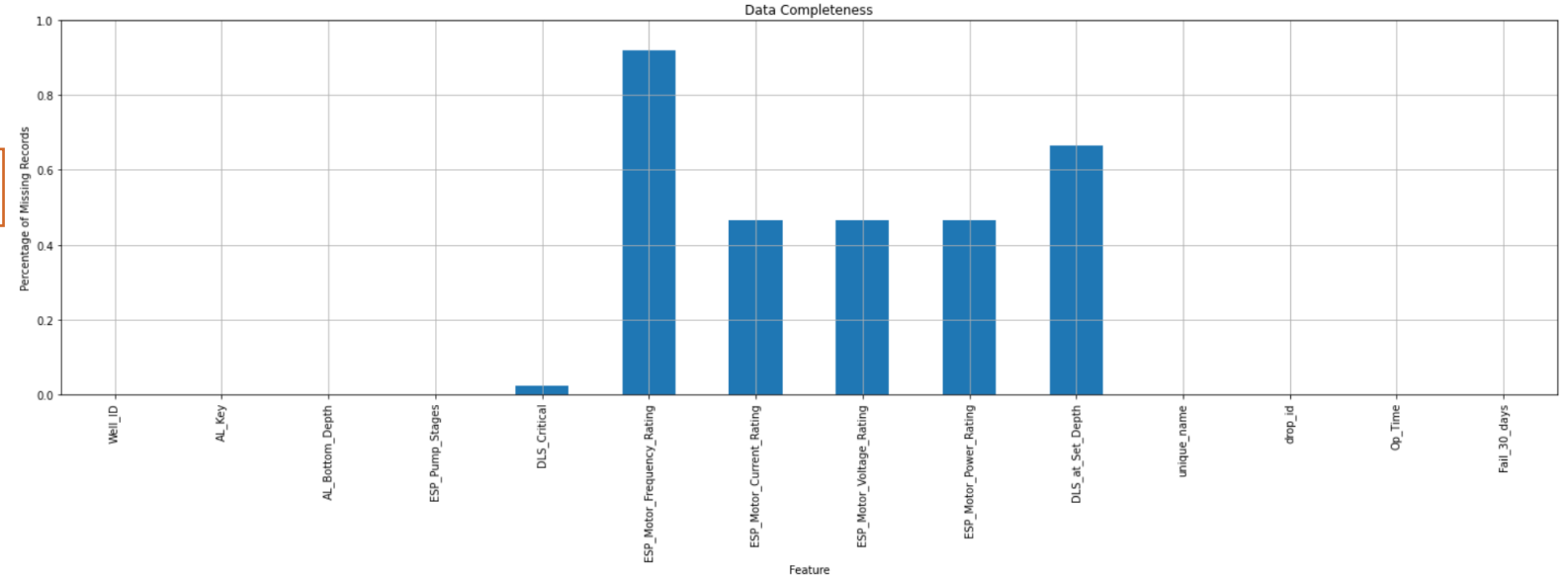


Feature
Imputation

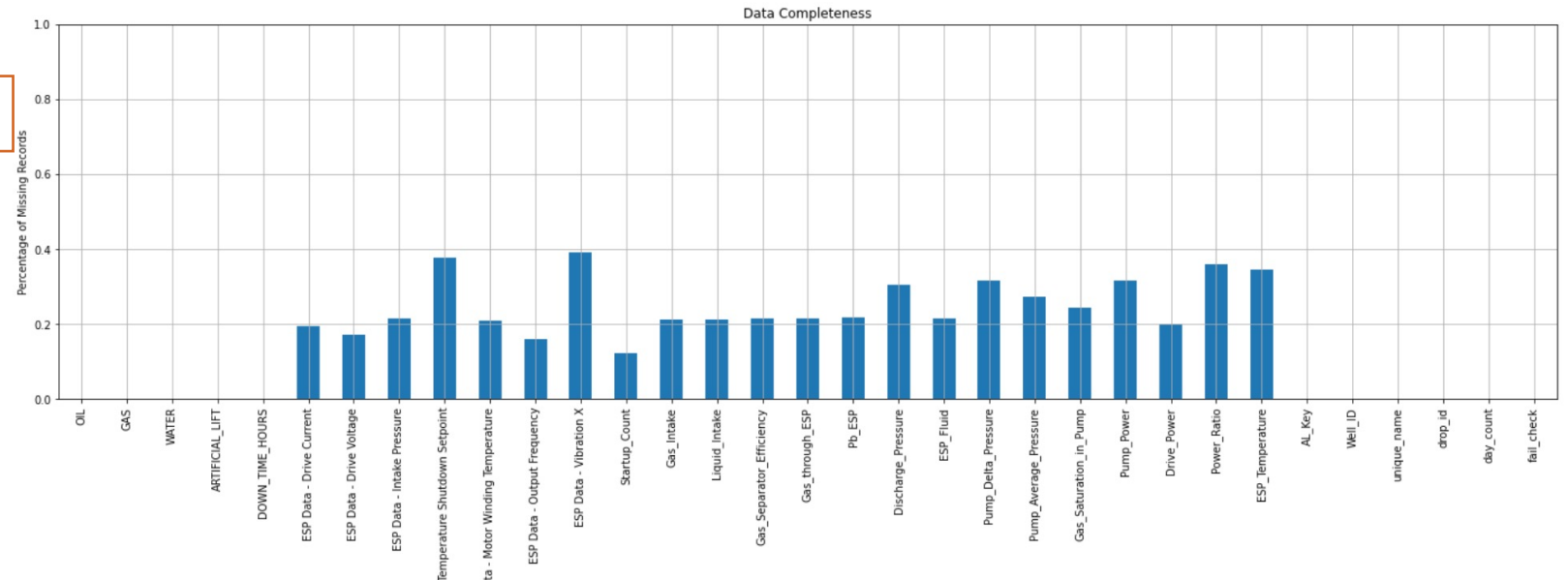


Feature
Ranking

Well Data



Daily Data





The University of Texas at Austin

Hildebrand Department of Petroleum
and Geosystems Engineering

Cockrell School of Engineering

1. Likewise
Deletion

2. Replace
with Mean

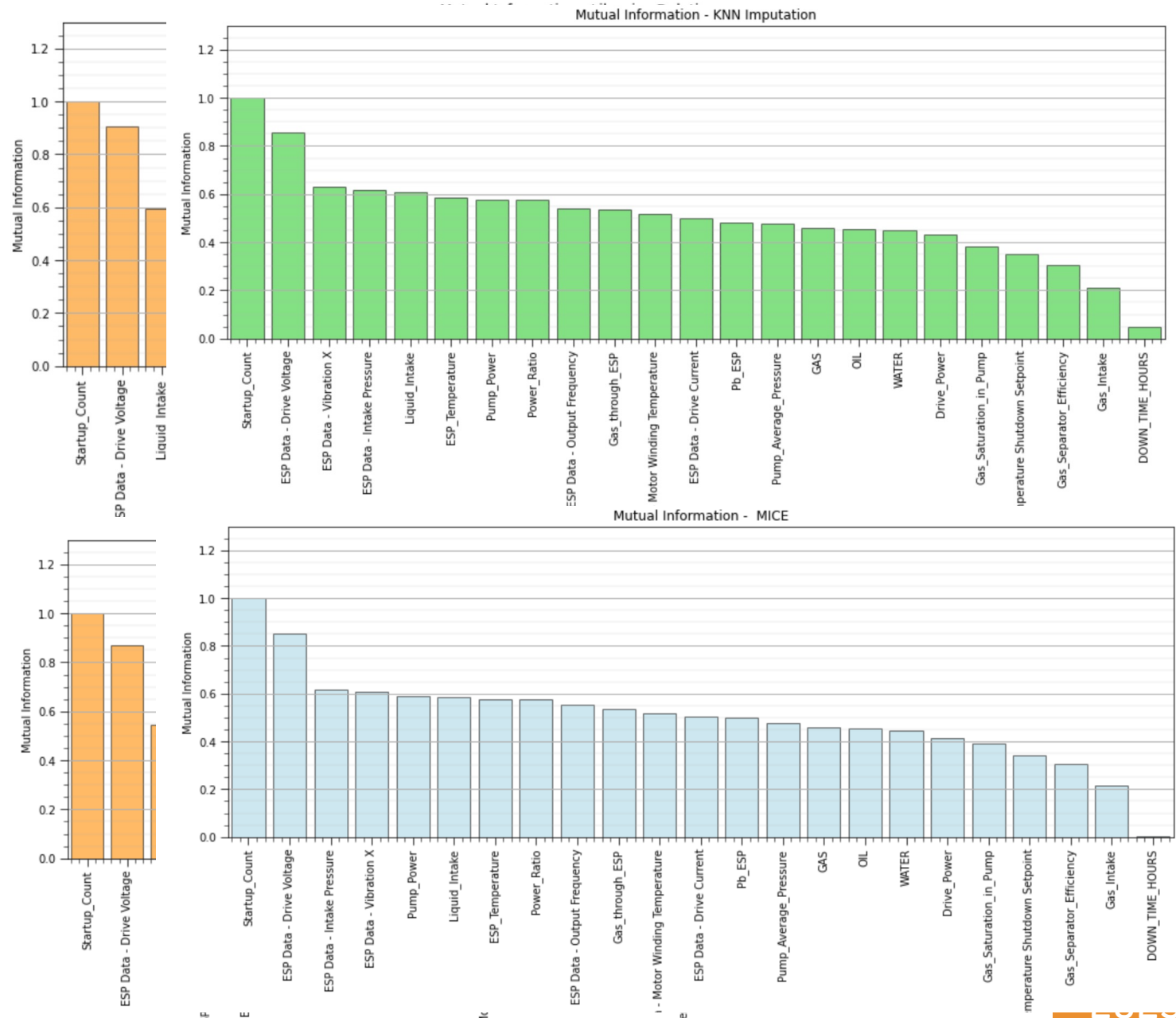
3. MICE

4. KNN

PCA

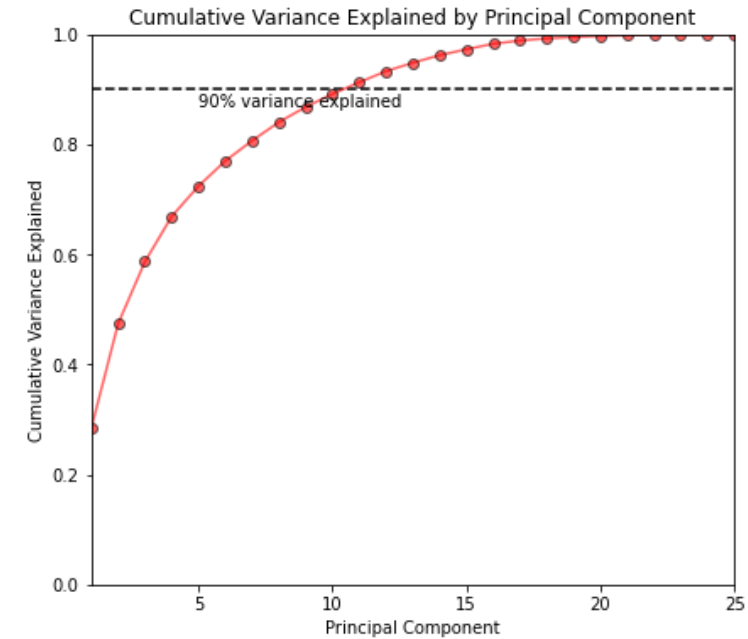
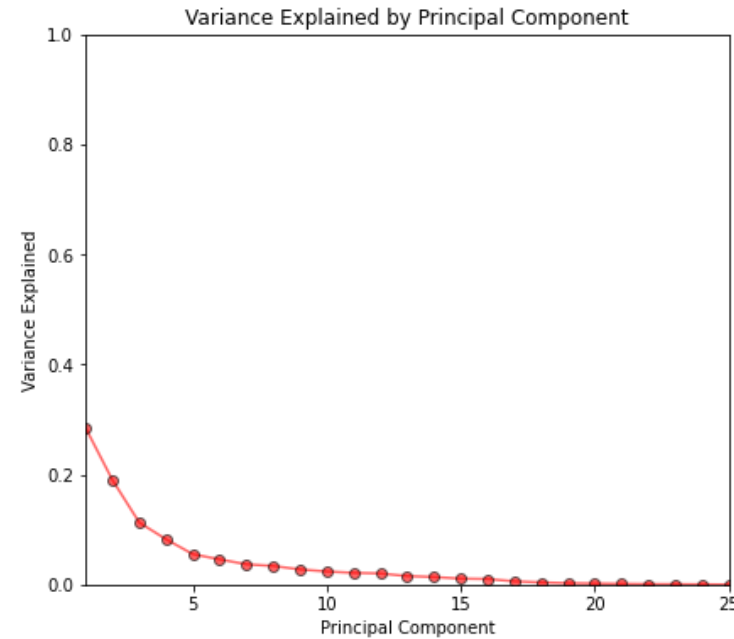
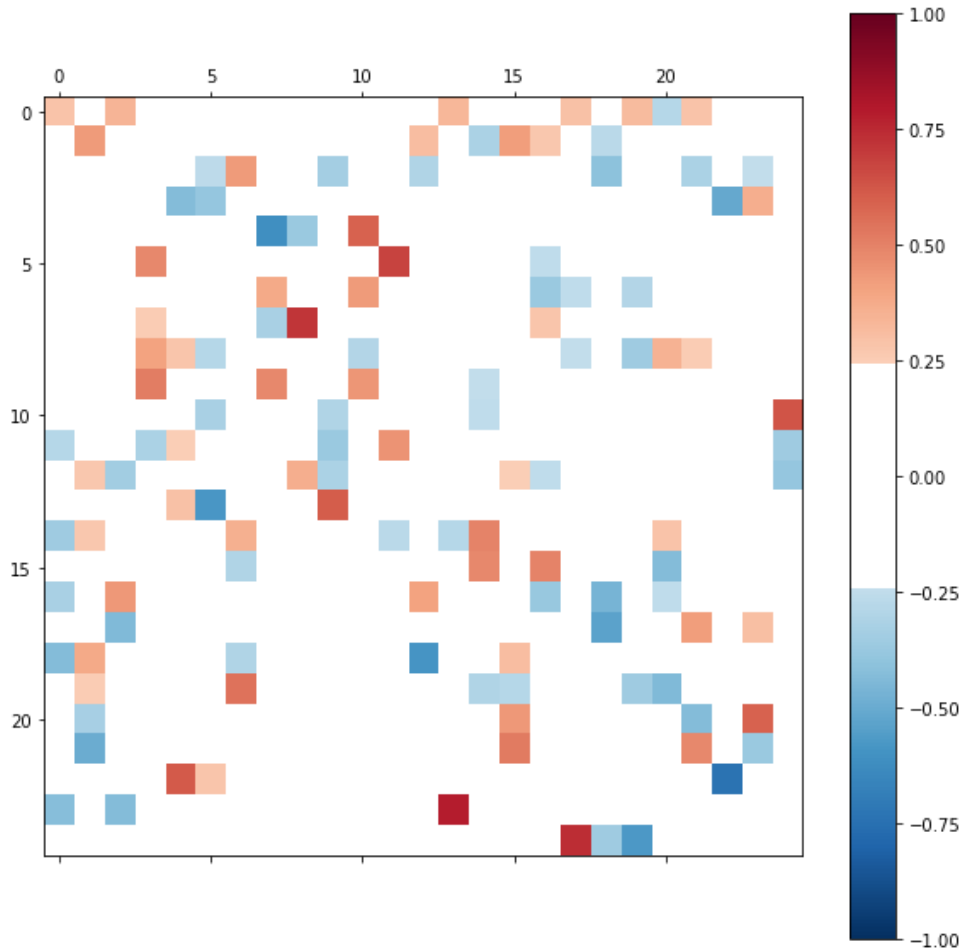
Mutual
Information

FEATURE IMPUTATION





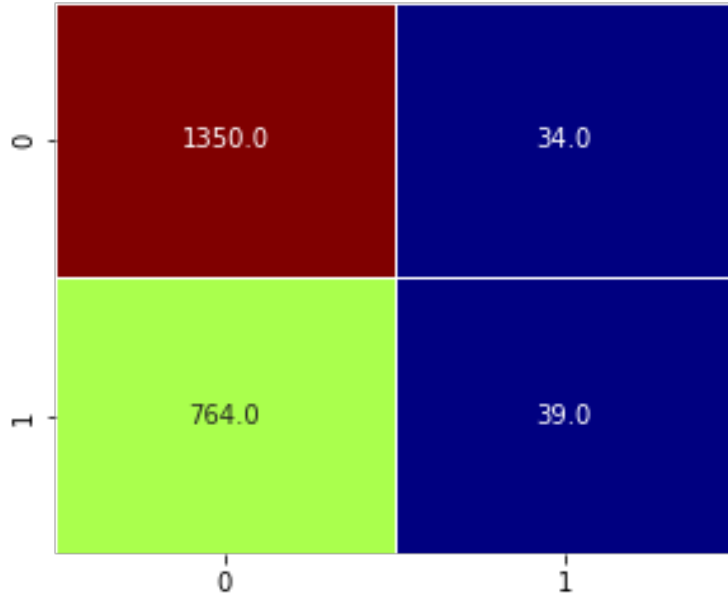
PCA Analysis





ML Models

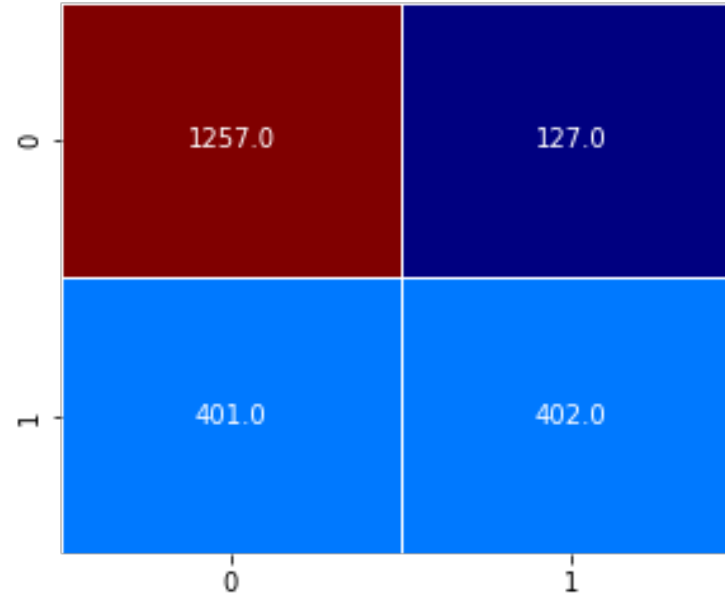
Logistic Regression Confusion Matrix



Accuracy

0.635

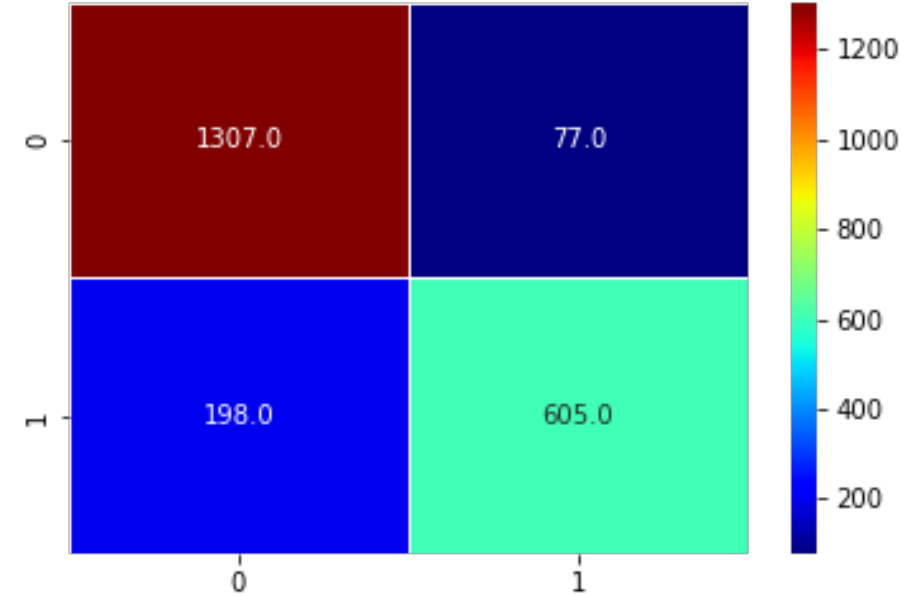
Decision Tree Confusion Matrix



Accuracy

0.759

Ensemble Tree Confusion Matrix



Accuracy

0.874





The University of Texas at Austin

Hildebrand Department of Petroleum
and Geosystems Engineering

Cockrell School of Engineering

Fail → **1**
No Fail → **0**

Well_ID	AL_Key	Fail in 30 days
327	ESP_1	1
125	ESP_2	0
260	ESP_3	0
21	ESP_1	1
16	ESP_1	0
20	ESP_1	0
76	ESP_1	0
60	ESP_1	1
13	ESP_1	0
10	ESP_2	1
27	ESP_1	0
105	ESP_2	0
53	ESP_2	1
84	ESP_1	0
12	ESP_1	1
113	ESP_2	0
62	ESP_1	0
25	ESP_1	0
74	ESP_3	1
7	ESP_2	0

Results

Well_ID	AL_Key	Fail in 30 days
91	ESP_1	1
17	ESP_1	1
132	ESP_5	0
116	ESP_1	0
90	ESP_1	0
61	ESP_1	0
101	ESP_1	0
6	ESP_1	0
93	ESP_1	1
248	ESP_2	0
123	ESP_2	0
32	ESP_1	0
269	ESP_1	1
129	ESP_1	0
139	ESP_1	1
254	ESP_1	1
253	ESP_1	0
95	ESP_1	0
258	ESP_1	0
44	ESP_1	1



The University of Texas at Austin

Hildebrand Department of Petroleum and Geosystems Engineering

Cockrell School of Engineering